A Linked Dataset of Medical Educational Resources

Editor(s): Name Surname, University, Country
Solicited review(s): Name Surname, University, Country
Open review(s): Name Surname, University, Country

Hong Qing Yu[a*], Stefan Dietze[b], Davide Taibi[c], Daniela Giordano[d], Eleni Kaldoudi[e] and John Domingue[a]
[a]*Knowledge Media Institute, The Open University, United Kingdom*
[b]*L3S Research Center, Leibniz University, Hanover, Germany*
[c]*Italian National Research Council, Institute for Educational Technologies, Italy*
[d]*University of Catania, Dipartimento di Ingegneria Elettrica, Elettronica e Informatica, Catania, Italy*
[e]*Democritus University of Thrace, Greece*

Abstract. Reusable educational resources became increasingly important for enhancing learning and teaching experiences, particularly in the medical domain where resources are particularly expensive to produce. With respect to this, research has aimed at improving interoperability across educational resources metadata repositories, which led to a fragmented landscape of competing metadata schemas, such as IEEE LOM or OAI-DC, and interface mechanisms, such as OAI-PMH or SQI. Thus, the major issue of educational resource federation is the heterogeneity challenge of metadata and data. In this paper, we illustrate a medical educational dataset (mEducator Linked Educational Resources dataset) that is published as part of the Linked Open Data cloud following Linked Data (LD) principles. The dataset contains educational resource metadata harvested from ten different (medical) educational institutes. In addition, the dataset has been improved and enriched by deploying a range of LD-based techniques and datasets. We introduce a Semantic Web Service-based data extracting mechanism that is exploited for services and data integration in order to address heterogeneous metadata problems together with deployed enrichment and clustering mechanisms. The paper also discusses the data accessibility via a set of dedicated APIs, use cases and applications, which exploit our dataset together with some statistics and evaluation results.

Keywords: Linked Data, medical educational resource, Semantic Web, Web services, Web APIs, learning repository, TEL

1 Introduction

Sharing and reusing educational resources has long been an overall vision in open distance education and particularly, within the open educational resources (OER) community. However, while interoperability is a fundamental requirement in resources sharing, the landscape of deployed technologies is still very fragmented [4]. This has led to vast amounts of educational resource metadata becoming available on the Web as part of still rather isolated and disparate educational resource metadata silos, such as ARIADNE[1], OpenLearn[2] or the different repositories of the OpenCourseWare[3] (OCW) Consortium. Whilst the Linked Data (LD) approach has emerged as de-facto standard for data sharing on the Web, recently we also started to witness increasing usage of LD [1] technologies for sharing of educational resources.

Since a large amount of educational data is already available on the Web via proprietary and/or competing schemas and interface mechanisms, the main challenges are to (a) start adopting LD principles and vocabularies while (b) leveraging on existing educational Web data accessible via non-LD compliant means and (c) improving interoperability at different levels, such as interfaces and schemas.

In this paper, we illustrate the mEducator Linked Educational Resources dataset that has been published in the Linked Open Data (LOD) cloud as one central outcome of the mEducator project[4]. By applying a Linked Services-based data federation framework [7], the mEducator dataset integrates medical educational resources from 10 heterogeneous educational repositories,

1 http://www.ariadne-eu.org/
2 http://www.open.edu/openlearn/
3 http://ocw.mit.edu/index.htm
4 http://www.meducator.net

by deploying mechanisms for data lifting from heterogeneous schemas and formats into a unified RDF schema as well as LD-based data enrichment and clustering techniques. The dataset is updated daily by dynamically harvesting the educational resource metadata through the repository Web APIs/Services.

2 The dataset: schema, storage and access

2.1. Educational resources sharing

Throughout the last decade, research in the field of technology-enhanced learning (TEL) has focused fundamentally on enabling interoperability and reuse of learning resources and data across the Web. Sharing and improving accessibility of educational resources from distinct resource collections and repositories increases the coverage and quality of resources offered by state-of-the-art information retrieval mechanisms and educational end-user applications. In addition, resource sharing reduces the costs of creating expensive medical educational objects such as virtual patients in the medical field, and lowers barriers for educational service providers. However, even though a variety of OER repositories, schemas and vocabularies [4] have been emerged throughout the last decade, the landscape is still highly fragmented and interoperability across disparate collections and federated, or even Web-wide search of OER is still an open challenge.

2.2. The mEducator resources schema

The strive for educational resource interoperability has led to a fragmented landscape of competing metadata schemas, i.e., general-purpose ones such as Dublin Core [3] or schemas specific to the educational field, like IEEE Learning Object Metadata (LOM) [2] or ADL SCORM[5] but also interface mechanisms such as OAI-PMH[6] or SQI[7]. To this end, although a vast amount of educational content and data are shared on the Web in an open way, the integration process is still costly as different learning repositories are isolated from each other and based on different implementation standards.

In order to define a suitable medical educational resource metadata-sharing schema, widely used learning object modelling standards (e.g. IEEE LOM and Dublin Core) and their adoption on the Web were studied. This led to the identification of frequently used concepts and properties as well as the detection of gaps with respect to the availability of suitable vocabularies. For instance, the description of disciplines, licensing schemes or learning outcomes. Based on the studies, the mEducator schema[18] was defined by integrating those commonly used concepts, vocabularies and properties under one unified model that was then extended by medical educational domain experts. Whilst aiming for a rather light-weight and usable, yet sufficiently comprehensive schema, the mEducator metadata schema[8] covers the most frequently used aspects of educational resources – from basic ones such as title and descriptions to more sophisticated ones such as learning outcomes and licensing models. In addition, a number of complementary vocabularies are provided.

2.3. Data storage and access

All extracted and generated educational metadata is eventually stored in a dedicated RDF store as part of the mEducator dataset.[9] The store later is implemented using Sesame/BIGOWLIM[10]. Each educational resource metadata entity described using the mEducator resource schema owns a unique and de-referencable URI, such as http://purl.org/meducator/resources/25a8c581-66d7-4186-9411-f9f0f783463e. In addition, a set of dedicated REST APIs is implemented to enable client applications to query, store and retrieve the metadata in the RDF store (an authentication key is required to access updating related APIs). These REST APIs facilitate different types of queries:

- SPARQL [19] query[11]
- Keyword-based query[12]
- property-based keyword query[13]
- rdfs:seeAlso-based keyword query[14]
- identifier-based property query[15]
- POST method for inserting an education resource into the RDF repository[16]

3 Data sources, generation and maintenance

The proposed architecture has been introduced in [20] and fully described in [21]. The data sources are originally

[5] Advanced Distributed Learning (ADL) SCORM: http://www.adlnet.org

[6] Open Archives Protocol for Metadata Harvesting: http://www.openarchives.org/OAI/openarchivesprotocol.html

[7] Simple Query Interface: http://www.cen-ltso.net/main.aspx?put=859

[8] http://www.purl.org/meducator/ns/

[9] http://ckan.net/package/meducator

[10] http://www.ontotext.com/owlim/

[11] http://meducator.open.ac.uk/resourcesrestapi/rest/meductor/auth/sparql?query=

[12] http://meducator.open.ac.uk/resourcesrestapi/rest/meducator/auth/keywordsearch?keyword=

[13] http://meducator.open.ac.uk/resourcesrestapi/rest/meductor/auth/propertysearch?property=&value=

[14] http://meducator.open.ac.uk/resourcesrestapi/rest/meductor/auth/eidsearch?id=

[15] http://meducator.open.ac.uk/resourcesrestapi/rest/meductor/auth/search?ids=&properties=

[16] http://meducator.open.ac.uk/resourcesrestapi/auth/rest/meducator/

derived from 10 medical educational metadata repositories (see Table 1). Based on these original metadata resources, we further conducted enrichment data via DBPedia spotlight[17] and BioPortal APIs[18] to enable linking our dataset to Linked Open Data Cloud and enhance the interlinks between resources.

Table 1 The list of data source repositories

Service	Base URI
Pubmed	http://www.pubmedcentral.nih.gov/oai/oai.cgi
University of Catania OER	http://151.97.9.184/WebTraceRDFApi/Default.aspx?
Biblioteca Digital de Teses e Dissertações da UFRN	http://bdtd.bczm.ufrn.br/tedesimplificado/tde_oai/oai2.php
Aristotle University of Thessaloniki, Moodle repository	http://kedip.med.auth.gr/meducatormoodleapi/index.php
Open Research Online (ORO), Open University, UK	http://libeprints.open.ac.uk/cgi/oai2
Medical learning resources, Technical University of Cluj-Napoca	http://dataserver.mediogrid.utcluj.ro/adnotare/cluj_endpoint.php
Educational resources of EUREKA project (Canada)	http://eureka.ntic.org/oai-pmh.php
Learning resources, University of Helsinki (Finland)	https://helda.helsinki.fi/dspace-oai/request
Linked Data Learning resources, University of Helsinki (Finland)	https://helda.helsinki.fi/mEducatorWebapp/RDFEndpoint
Nice University educational resources (France)	http://revel.unice.fr/oai/oai2.php

The educational repositories are integrated through annotation, dynamically invoking and lifting the responses of the Web APIs. Three steps facilitate our data extraction process: (a) Manually annotating Web APIs/services using Linked Data technologies, (b) schema mapping and (c) data lifting. First of all, a service requester (or provider) needs to manually provide a mapping schema as part of step (a). This template defines the mapping of the source schema to the desired RDF target schema. Then our service invocation engine applies the mapping schema to lift the original service response to the correct RDF description as part of step (b) and (c).

To achieve this goal, we exploit two well-integrated technologies: iServe[19] [7] and SmartLink[20] [9]. iServe provides a framework for LD-based service modelling and supports related tools for dynamic discovering and invoking Web APIs. SmartLink is a LD-based and iServe compliant Web tool[21] that handles two different kinds of service semantic annotations separately, namely functional and non-functional service annotations. The service requester able specifies the mapping schema between Web API responses and the requester desired RDF output formats using the XSPARQL language[22]. The service invocation engine (OmniVoke) [16] dynamically lifts the Web API output to the required RDF (in our case, it is an mEducator schema based RDF presentation). In detail, the 3-step process of resources extraction is presented in Figure **1** .

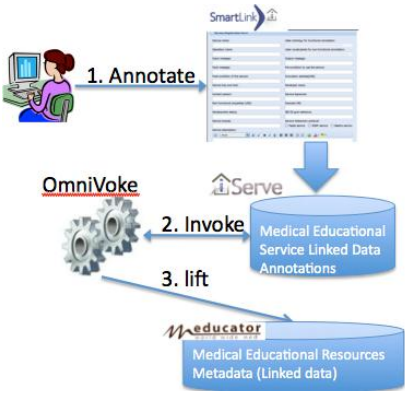

Figure 1 Medical educational resources metadata extraction process

With respect to the services integrated into the particular dataset described here, we would like to highlight that even though only a fraction of the source metadata schemas is mapped into our RDF schema, the lifted metadata instances still contain the vast majority of metadata. This is due to the fact that in most cases, metadata records found on the Web use only a small subset of the otherwise rather complex educational resource schemas. To this extent, we have deliberately chosen to only reflect metadata properties that are widely used by educational resource providers on the Web.

4 Data enrichment and clustering

4.1. Data enrichment for interlinking with external datasets

Though metadata is lifted into RDF automatically, it is often poorly structured and makes only very limited and fragmented use of controlled vocabularies. To alleviate this problem, we further enrich the resource metadata by taking advantage of available LOD datasets such as DBpedia (via DBpedia Spotlight) and the multitude of ontologies available via the BioPortal APIs. These enrichments are utilised for (a) expanding existing metadata with publicly available knowledge, (b) disambiguation of data and (c) clustering correlated resources by exploiting the use of shared vocabularies.

Data enrichment is implemented in two ways (a) as an automated mechanism whenever new data is pushed to the RDF store and (b) as semi-automated approach that

[17] http://dbpedia.org/spotlight

[18] http://www.bioontology.org/wiki/index.php/BioPortal_REST_services

[19] http://iserve.kmi.open.ac.uk/

[20] http://smartlink.open.ac.uk

[21] http://smartlink.open.ac.uk/smartlink

[22] http://xsparql.deri.org/

provides end-users with suggestions of related entities that match a particular term. A user can select suitable ones as part of a particular end user application. While the first approach makes usage of DBpedia exclusively, resulting in large numbers of automatically retrieved references to DBpedia resources, the second approach makes exclusive use of the BioPortal API which provides access to over 300 bio-medical ontologies and 5 million entities. Currently, we target a small number of usually poorly structured properties with this process (titles, keywords, descriptions, subjects, disciplines). These were selected because (a) they cover essential information required by recommendation systems and (b) educational resources found on the Web usually describe such properties by means of unstructured text and highly heterogeneous taxonomies, while the use of structured (domain-specific and cross-domain) vocabularies is highly recommended to improve interoperability across different datasets.

Table 2 Distribution of the sources of the used terms.

External Source	Number of distinct Terms	Percentage
DBPedia	509	93.39
Medical Subject Headings	11	2.02
SNOMED Clinical Terms	11	2.02
Health Level Seven	4	0.73
Galen	2	0.37
MedDRA	2	0.37
LOINC	1	0.18
MedlinePlus Health Topics	1	0.18

The number of added enrichment triples in the data store is 1352. Table 2 shows that the added enrichment triples involve a total of 509 distinct terms from DBpedia. The average number of enrichments per enriched resource is 4.5 (min=1, max=42). Apparently, a large number of enrichments are obtained via the automated enrichment based on the DBpedia Spotlight API, while the semi-automated approach from the BioPortal API provides a higher diversity (data from different vocabularies such as MESH and SNOMED are used). However, only a very limited amount of overall enrichments are obtained because it requires manual intervention and pre-selection of suggested terms. Figure 2 shows an example of enrichment triples that added links to the DBpedia and BioPortal terms for explaining "thrombolysis", which describes the subject of a medical educational resource.

Figure 2 An example of added enrichment triples.

4.2. Data clustering for internal interlinking

Clustering functionalities have been integrated into the RDF store to allow the interlinking of related resources originating from different repositories. We perform two forms of clustering: (i) clustering based on linguistic metadata similarity and (ii) structural clustering. The clustering process for (i) involves three main steps (see [21]): content indexing, creation of a similarity matrix and clustering. The first step parses the RDF fields of the selected subsets of the unifying metadata schema, including those that contain descriptive free text, and creates a matrix, namely: Doc-Term (DT). The DT matrix contains the frequency of each term in a given resource and consolidates information about term co-occurrences within the same resource and across all resources, which is done according to a method that weights less common terms and also takes the neighboring terms into account as term context [8]. The second step is based on the DT matrix to create the similarity matrix(S) by assigning a similarity score across the resources. The S matrix is then used to classify the resources by applying a clustering algorithm. Our implementation supports both clustering based on Kohonen maps and Aggregative clustering.

With respect to (ii) we exploited the enrichments generated in the previous steps to detect correlations between resources. Therefore, we created different views on the generated resources graph by clustering resources according to different thresholds(T), where T defines the minimum amount of equivalent enrichments of all resources R within a particular cluster (C). For instance, applying a threshold T=2 led to 13 clusters in which all resources shared at least 2 enrichment concepts. In contrast, lower (7) but more coherent clusters were created using T=4, indicating that there will be an edge between two nodes only if they have more than four concepts in common.

The following table shows the nodes generating by using a cluster of the graph generated with T=2. Manual assessment proved a strong semantic correlation of the clustered resources, i.e. the majority of the enrichment concepts used in this cluster are related to similar diseases or disease aspects (arteries, vascular, ischemia, aneurysm, atherosclerosis).

Table 3 Cluster 2 enrichments/resources (T=2)

	infection	atherosclerosis	syndrome	reading	evaluation	arteries	vascular	ischemia	aneurysm	symptoms
406eb475-67ae-49ad-848b-0a166a59389c	0	0	0	1	1	0	0	1	0	0
e72b12bd-87ae-4035-9799-0ed2a4751181	1	1	0	1	1	1	1	0	1	1
1928266c-2e8f-477f-858e-085b0eb55555	0	0	1	1	1	0	0	0	0	1
2a5360c9-8284-4dc2-8534-3544fd7e67c2	0	1	0	1	1	0	0	0	0	0
257093d6-b1d1-4016-b131-4d7ca11dac07	1	0	1	1	1	1	0	1	0	0
649e2272-1e66-4fb2-8073-2dfe9d0b59f1	0	0	0	1	1	0	0	0	0	0
8b199664-042a-4974-9716-6ed15b3bda81	0	0	0	1	1	0	1	1	1	0

Similar results were obtained with clusters generated with T=4, whilst even less noise was detected.

To treat the clustering results consistently with the LD approach and integrate this information in the unified mEducator metadata schema, a separate RDF schema has been defined (see Figure 3 for an excerpt of the relative instance). In this schema, the classes and properties are defined to fully describe not only the clustering results, but also the complete rundown of the clustering process itself. It adopts the Provenance Vocabulary[23] that contains all the necessary properties and classes to describe the provenance of data from the Web. In the clustering schema, the data includes, but is not limited to, the clustering processing initiator agent (human, or software), the features used for indexing, the algorithm used and its parameters, the linked resources and clusters.

```xml
<mcc:Cluster rdf:about="http://purl.org/meducator/cluster/3728724799kjd92jdd020">
  <rdfs:label>Cluster 3728724799kjd92jdd020</rdfs:label>
  <mcc:identfifier>3728724799kjd92jdd020</mcc:identfifier>
  <mcc:clusterType>manual</mcc:clusterType>
  <mcc:containsResource rdf:resource="http://purl.org/meducator/resource/7284dfhje398udsuf93fujs"/>
  <mcc:containsResource rdf:resource="http://purl.org/meducator/resource/7284dfhje398udsuf93fuj2"/>
  <mcc:containsResource rdf:resource="http://purl.org/meducator/resource/7284dfhje398udsuf93fuj3"/>
  <mcc:containsResource rdf:resource="http://purl.org/meducator/resource/7284dfhje398udsuf93fuj4"/>
  <mdc:description>this could be any predefined id if available</mdc:description>
  <prv:createdBy>
    <prv:DataCreation>
      <prv:employedArtifact>
        <mcc:Algorithm rdf:about="http://www.purl.org/meducator/clustering/ReferenceAlgorithm">
          <mcc:hasConfidenceLevel>0.6</mcc:hasConfidenceLevel>
          <mcc:type>manual</mcc:type>
          <mcc:usesFeature>mdc:title</mcc:usesFeature>
          <mcc:usesFeature>mdc:description</mcc:usesFeature>
          <mcc:algorithmDescri
        </mcc:Algorithm>
      </prv:employedArtifact>
    </prv:DataCreation>
  </prv:createdBy>
  <mcc:hasCommonFeatures>
    <mcc:CommonFeatures>
      <mcc:title>Thrombolysis</mcc:title>
      <mcc:title>ECG</mcc:title>
    </mcc:CommonFeatures>
  </mcc:hasCommonFeatures>
</mcc:Cluster>
```

```xml
<mdc:Resource rdf:about="http://purl.org/meducator/resource/7284dfhje398udsuf93fujs">
  <mdc:identifier>http://www.med.helsinki.fi/tuke/meducator/Kansio 1004/Patient%201004.pdf</mdc:identifier>
  <mdc:title><![CDATA[ ECG Patient case 1005 limb and chest leads & thrombolysis ]]></mdc:title>
  <mcc:belongsToCluster rdf:resource="http://purl.org/meducator/cluster/3728724799kjd92jdd020"/>
  <mdc:rights rdf:resource="http://purl.org/meducator/licenses/Attribution-Non-Commercial-Share-Alike"/>
  <mdc:metadataLanguage>English</mdc:metadataLanguage>
  <mdc:creator>
```

Figure 3 An RDF instance of a cluster (large bounding box) and the relative reference in the mEducator resource's instance (small bounding box).

The clustering process receives input mEducator resources from all the subscribed repositories in the RDF/XML format. The process initiator selects the desired fields that should be taken in consideration in the clustering. The output of the system is another RDF/XML file that is uploaded to all queried repositories.

5 Data Usages

Currently, four educational Web applications[24] have been developed that interact with the mEducator Linked dataset. For example, the data and services integration APIs and datasets presented in the previous sections are fully integrated in the Metamorphosis+ [22] environment, which merges the paradigms of semantic and social web for sharing linked educational resources. MetaMorphosis+ realises the educational application and presentation layer. It allows viewing, management and annotation of the educational resource metadata that is retrieved via the APIs provided by the educational services and data integration mechanisms introduced here. Meanwhile, additional medical educational applications have emerged which interact with our mEducator dataset, which are MILES+[25], MELINA+[26] and LINKED LABRINTH[27].

6 Statistical Analysis

By the 1st May 2012, the mEducator Linked Educational Resources RDF repository contained in total 23876 triples, of which 10206 directly refered to a total 375 distinct educational resources. The average number of triples per educational resource is 27, ranging from a minimum of 6 to a maximum of 68. Figure **4** provides an overview of the heterogeneity of the origin of individual resource metadata.

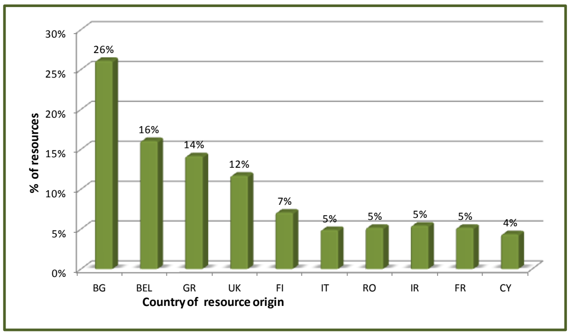

Figure 4 Number of resources (%) per country of origin (based on country of origin of contributing institution).

In addition, the metadata imported from external stores is usually very limited and it often covers only few properties, e.g. title, description and resource location. Based on our automated and semi-automated enrichment techniques, richer descriptions that populate additional properties are provided for the majority of resources, resulting all resources having a minimum of 5 described properties.

The usage frequency of particular properties in our RDF schema reflects the importance of particular schema elements (e.g. when considering the source metadata descriptions) and the richness of the transformed and enriched resources descriptions (when considering the improved metadata in our dataset). Table **4** provides an overview of property usage frequency across educational resources. Properties are listed in the order of decreasing frequency. A more detailed description of individual properties is given in [18]. Data collected in the RDF store makes use of a set of external LOD vocabularies/datasets. Table **4** shows the distribution of the terms used by external sources, and the percentage of usage of each source.

Table 4 mEducator schema property usage frequency across educational resources in the dataset

Property	%	Property	%
mdc:creator	100	mdc:hasEnrichmentContext	79.2
mdc:description	100	mdc:rights	77.9
mdc:metadataCreated	100	mdc:discipline	76.5
mdc:metadataCreator	100	mdc:technicalDescription	68.3
mdc:title	100	mdc:educationalContext	66.7
mdc:created	95.7	mdc:educationalLevel	64
mdc:metadataLanguage	95.5	mdc:educationalObjectives	58.9
mdc:citation	94.4	mdc:teachingLearningInstructions	57.9
mdc:language	89	mdc:educationalOutcomes	51.2
mdc:resourceType	88.6	mdc:educationalPrerequisites	49.3
		mdc:disciplineSpeciality	45.6
		mdc:assessmentMethods	40.8

[23] http://trdf.sourceforge.net/provenance/ns.html

[24] http://www.meducator3.net/

[25] http://kedip.med.auth.gr/meducator3/milesplus/

[26] http://www.meducator3.net/melinaplus/

[27] http://www.meducator3.net/llplus/

Table **5** lists the properties used to trigger the enrichment of the resources and the number of enrichments obtained. It is noted that only the three free text properties are currently considered during the enrichment process.

Table 5 Distribution of external DBpedia terms to resource properties

Enriched Property	Number of enrichments	Percentage
mdc:description	733	54.2
mdc:title	327	24.2
mdc: educationalOutcomes	292	21.6

7 Qualitative evaluation of enrichments

To evaluate the quality of the automated enrichment procedure results, a set of 200 enrichments from the overall set of 1352 has been selected randomly and assessed manually by experts. Based on their assessment, the experts assigned one of 3 possible ratings to each of the enrichment to indicate its grade of meaningfulness:

(A) The description of the external enrichment concept describes and expands the semantic meaning of the enriched source text.

(B) The enrichment concept shows a significant relationship with the enriched text but does not provide a direct mapping. This category also was assigned to enrichments where the correctness could not be determined with certainty due to ambiguity of the source concept.

(C) The enrichment has no meaningful relationship with the source concept.

This evaluation led the following results: 92% of enrichments belong to the category (A), while the 8% of enrichments are equally distributed between the categories (B) and (C).

Table 6 Assessment results

Category	n# enrichments	%
A	184	92
B	8	4
C	8	4

Assigning a value of 1 for the category A, 0.5 for the category B and 0 for the category C the overall average score for the selected enrichments is: 0.94. While hypotheses H2 could be confirmed by the evaluation results, we also identified some minor amount of noise contributed by the current enrichment approaches. We are currently investigating ways to reduce the number of less accurate enrichments by, for instance, expanding the amount of considered context used by the enrichment queries and by replacing DBpedia Spotlight with typed queries to DBpedia.

Whilst the analysis did not yet assess in a structured manner the correctness of the generated clusters (H3), it has confirmed its potential for disambiguation by associating acronyms and synonyms with enriched text. For example, description of medical learning resources showing the ambiguous acronym HPV have been correctly enriched with references to http://dbpedia.org/resource/Human_papillomavirus based on the co-occuring terms in the resource description. Similar behavior has been detected with the acronyms TMJ (Temporomandibular joint) and EASA (European Aviation Safety Agency). Concerning synonyms, description containing the words: ‘pap’ ‘smear’ or ‘gastric cancer’ have been respectively enriched with corresponding pap_test and stomach_cancer concepts within DBpedia. Such kind of enrichments allows broader yet more precise search and clustering results (H3) which will be subject to further investigation.

8 Conclusion and future work

In this paper, we introduced the mEducator Linked Educational Resources dataset. The dataset is published using Linked Data principles to support Web-scale interoperability between educational resources. That is, educational services as well as data. Linked Data is adopted to describe both services and data, which allows the integration of existing educational repositories at both the service and the data levels. By exposing educational resources via Linked Data principles, we leverage the wealth of existing datasets and vocabularies, so that internal links between educational data and resources are generated. We have introduced a set of implemented integration approaches, resulting RDF datasets, APIs, and applications (e.g. MetaMorphosis+ and an Android app) that provide open environments for medical education. One of the longer-term goals is to use the principles described in this paper to establish a unified entry point to well-interlinked educational datasets on the Web.

Future work will focus on two major areas, including the investigation of methods to enable the integration of data from other educational domains; and the extension of the framework with additional open repositories and data stores to further showcase and evaluate our services and data integration approach.

Acknowledgments. This work is partly funded by the mEducator project (Contract Nr: ECP 2008 EDU 418006 mEducator) under the eContentplus programme of the European Commission.

References

[1] Bizer, C., T. Heath, et al. (2009). Linked data - The Story So Far. Special Issue on Linked data, International Journal on Semantic Web and Information Systems (IJSWIS).

[2] IEEE Learning Technology Standards Committee (LTSC). Learning object metadata. In IEEE Learning Technologies Standards Committee 2011. IEEE, 2011.

[3] Dublin core metadata element set, version 1.1, 2011.

[4] Dietze, S., Sanchez-Alonso, S., Ebner, H., Yu, H., Giordano, D., Marenzi, I., Pereira Nunes, B. (2013) Interlinking educational Resources and the Web of Data – a Survey of Challenges and Approaches, accepted for publication in Emerald Program: electronic Library and Information Systems, Volume 47, Issue 1 (2013).

[5] MedBiquitous Consortium. Healthcare lom overview. In MedBiquitous Consortium 2011, 2011.

[6] Mitsopoulou Evangelia, Davide Taibi, Daniela Giordano, Stefan Dietze, Hong Qing Yu, Panagiotis Bamidis, Charalampos Bratsas, and Luke Woodham. Connecting medical educational resources to the linked data cloud: the meducator rdf schema, store and api. 1st International Workshop on eLearning Approaches for the Linked Data Age (LinkedLearning2011), 2011.

[7] Pedrinaci, C., Liu, D., Maleshkova, M., Lambert, D., Kopecky, J., and Domingue, J. (2010) iServe: a Linked Services Publishing Platform, Workshop: Ontology Repositories and Editors for the Semantic Web at 7th Extended Semantic Web Conference.

[8] Cohen, T., Schvanevendt, R., , Widdows, D. Reflective Random Indexing and indirect inference: a scalble method for discovery of implicit connections. J Biomed Inform. 43(2):240-56. 2010

[9] Dietze, S., Yu, H.Q., Pedrinaci, C., Liu, D. and Domingue, J. (2011) SmartLink: a Web-based editor and search environment for Linked Services, 8th Extended Semantic Web Conference (ESWC), Heraklion, Greece

[10] Hadzic, M., D'Souza, R., Hadzic, F., Dillon, T. Thinking PubMed: an Innovative System for Mental Health Domain. Computer-Based Medical Systems, 21st IEEE International Symposium on Computer-Based Medical Systems, 2008

[11] Henze, N. (2006). Personalized E-Learning in the Semantic Web. Extended version of 4. International Journal of Emerging Technologies in Learning (iJET), 1(1).

[12] Henze, N., Dolog, P., and Nejdl, W. (2004). Reasoning and Ontologies for Personalized E-Learning. Educational Technology & Society, 7(4).

[13] IEEE, IEEE Standard for Learning Object Metadata, IEEE Std 1484.12.1-2002, vol., no., pp.i-32, 2002, doi: 10.1109/IEEESTD.2002.94128.

[14] Kaldoudi E, Dovrolis N, Giordano D, Dietze S., Educational Resources as Social Objects in Semantic Social Networks, Proceedings of the 1st International Workshop on eLearning Approaches for the Linked Data Age, Herakleio, May 2011

[15] Kobilarov, G. and Dickinson, I. (2008). Humboldt: Exploring Linked Data, Proc. Linked Data on the Web (LDW'08), Beijing, China, April 2008

[16] Li, N., Pedrinaci, C., Maleshkova, M., Kopecky, J. and Domingue, J. (2011) OmniVoke:A Framework for Automating the Invocation of Web APIs, Fifth IEEE International Conference on Semantic Computing, Stanford University, Palo Alto, CA, USA.

[17] Marchionini, G. (2006) Exploratory search: from finding to understanding. Commun. ACM, 49(4):41-46,

[18] Mitsopoulou, E., Taibi, D., Giordano, D., Dietze, S., Yu, H. Q., Bamidis, P., Bratsas, C., Woodham, L., Connecting Medical Educational Resources to the Linked Data Cloud: the mEducator RDF Schema, Store and API, Proceedings the 1st International Workshop on eLearning Approaches for the Linked Data Age, CEUR-Vol 717, 2011.

[19] World Wide Web Consortium, W3C Recommendation, SPARQL query language for RDF, 2008, (http://www.w3.org/TR/rdf-sparql-query/).

[20] Yu, H. Q., Dietze, S., Li, N., Pedrinaci, C., Taibi, D., Dovrolis, N., Stefanut, T., Kaldoudi, E., Domingue, J., A Linked Data-driven & Service-oriented Architecture for Sharing Educational Resources, in Proceedings of the 1st International Workshop on eLearning Approaches for the Linked Data Age, CEUR-Vol 717, 2011.

[21] Dietze, S., Yu, H. Q., Giordano, D., Kaldoudi, E., Dovrolis, N., Taibi, D., Linked Education: interlinking educational Resources and the Web of Data, in Proceedings of the 27th ACM Symposium On Applied Computing (SAC-2012), Special Track on Semantic Web and Applications, Riva del Garda (Trento), Italy, 2012.

[22] Dietze, S., Kaldoudi, E., Dovrolis, N., Yu, H.Q., Taibi, D. (2011) MetaMorphosis+ – A social network of educational Web resources based on semantic integration of services and data, 10th International Semantic Web Conference (ISWC2011), Bonn, Germany.